

# A physical and genetic map of the *Spiroplasma citri* genome

Fengchun Ye<sup>1</sup>, Frédéric Laigret<sup>1\*</sup>, Jane C. Whitley<sup>1,2</sup>, Christine Citti<sup>1</sup>, Lloyd R. Finch<sup>2</sup>, Patricia Carle<sup>1</sup>, Joël Renaudin<sup>1</sup> and Joseph-Marie Bové<sup>1</sup>

<sup>1</sup>Laboratoire de Biologie cellulaire et moléculaire, 33883 Villenave d'Ornon Cedex, France and

<sup>2</sup>Russel Grimwade School of Biochemistry, University of Melbourne, Parkville, Victoria, Australia

Received January 17, 1992; Revised and Accepted March 3, 1992

## ABSTRACT

A physical and genetic map of the *Spiroplasma citri* genome has been constructed using several restriction enzymes and pulsed field gel electrophoresis. A number of genes were subsequently localized on the map by the use of appropriate probes. The genome size of the spiroplasma estimated from restriction fragments is close to 1780 kbp, the largest of all Mollicutes studied so far. It contains multisite insertions of *Spiroplasma virus 1* (SpV1) sequences. The physical and genetic map of the *S. citri* genome shares several features with that of other Mollicutes, especially those in the *Mycoplasma mycoides* cluster. This supports the finding that *S. citri* and these *Mycoplasma* spp. are phylogenetically related.

## INTRODUCTION

Organisms in the class Mollicutes are among the simplest cellular organisms known to date. These wall-less eubacteria are characterized by the small size of their genomes and a low G+C base content. There are six genera: *Mycoplasma*, *Ureaplasma*, *Spiroplasma*, *Anaeroplasm*, *Acholeplasma* and *Asteroleplasma*. Recent studies have revealed that the genome sizes of different species of *Mycoplasma* and *Ureaplasma* were heterogeneous and, consequently, genome sizes could not be used anymore as a criterion for taxonomic classification of these organisms (1–4).

The genome of *S. citri* is about 1780 kbp in size; it is the largest of those tested so far. *Mycoplasma genitalium* has the smallest genome (600 kbp) (5,6). Little is known about the genomic organization of *S. citri*. Comparisons of 16S and 5S ribosomal RNA (rRNAs) sequences from different Mollicutes indicated that *S. citri* is phylogenetically related to *Mycoplasma mycoides* and relatives (7). Because physical and genetic maps have been already obtained for several mollicutes (5,6,8–13) and especially for *M. mycoides* and relatives (8,13) it seemed of great interest to establish such map for the *S. citri* genome, in order to compare their respective genomic features. The result of this work is presented here.

## MATERIALS AND METHODS

### DNA preparation; separation of restriction fragments resulting from single digests

*S. citri* (strain R8H2. HP, ATCC=27556<sup>T</sup>) was grown at 32°C in BSR medium (14). Genomic DNA embedded in LMP agarose (Geneline<sup>TM</sup>, Beckman, Palo Alto, CA) was prepared as previously described (15), except that the cellular pellet was washed with STE (100mM NaCl, 10mM Tris-HCl pH 7.5, 1mM EDTA) instead of PBS buffer. Restriction enzymes *Apa*I, *Bss*III, *Not*I, *Sal*I, *Sma*I and *Sst*II were used as indicated by the manufacturers (GIBCO/BRL, Gaithersburg, MD; Biolabs, Beverly, MA) and digestions were performed for at least 6 hours. The reactions were stopped by replacing the digestion buffer with sterilized electrophoresis buffer. The restriction fragments were separated onto a 1% LE agarose gel (Geneline<sup>TM</sup>, Beckman) using a TAFE (Transverse Alternating Field Electrophoresis) apparatus (Beckman Instruments) and TAFE buffer recommended by the supplier. The separation conditions were as follows: Pulse times: 1s, 2hrs; 5s, 2hrs; 10s, 2hrs; 15s, 2hrs; 20s, 2hrs; 30s, 2hrs; 40s, 2hrs; 45s, 2hrs; 50s, 2hrs; 1min, 2hrs; 1min 20s, 2hrs; 5s, 45min; 15s, 45min; 30s, 45min; 45s, 45min at a voltage of 220V and a temperature of 8°C.

The size of all separated fragments was evaluated by comparison with yeast chromosomal DNA (BioRad, Richmond, CA) and lambda phage DNA concatemers as molecular weight markers.

### Linkage of the restriction fragments and physical mapping of the *S. citri* genome

The order, on the genome, of the restriction fragments produced by each restriction enzyme was established according to the method of Bautsch (10), using sequential double digestions with pairs of enzymes. However, some small fragments were difficult to map by this approach and, in this case, we used partial digestions and hybridizations as described in the results.

### Localization of genes on the physical map of *S. citri* genome

Gene markers, used as probes, are listed in Table 1. The plasmids in which they were inserted were propagated in *E. coli* strains DH5αF' or HB101 grown in LB medium supplemented with

\* To whom correspondence should be addressed

appropriate antibiotics. Inserts were separated from plasmid DNA by digestion with appropriate endonucleases followed by preparative agarose gel electrophoresis and elution using the Gene Clean kit (Bio101, La Jolla, CA).

TAFE gels were treated twice for 10 minutes with 0.25N HCl before transfer to nylon membrane (Hybond N+, Amersham, Buckinghamshire, UK) according to the standard alkaline Southern blotting procedure (26,27). The membranes were subsequently hybridized with one of the probes labelled with dCTP or dATP  $\alpha$ -<sup>32</sup>P using random priming kits (GIBCO/BRL). Total *S. citri* tRNAs were extracted according to the procedure described by Zubay (28) and labelled using T4 polynucleotide kinase (Biolabs).

The probe for *rpoD*, encoding the general Sigma factor, was obtained as follows. Two oligonucleotide primers were deduced from conserved amino-acid sequences in Sigma factors of different bacteria, as previously described for *Chlamydia trachomatis* (29). Their sequences are 5'-GG(A,T)(T,C)T(A,T)ATGAA(A,G)GC(A,T)GT-3' for the first primer and 5'-GC(T,C)TG(A,G,C,T)C(G,T)(A,G,T)AT(C,T)CA(C,T)CA(A,T)G-3' for the second primer. Polymerase chain reaction (PCR) was carried out in a DNA thermal cycler (Perkin-Helmer Cetus, Emeryville, CA) using 35 cycles, each of 2 min at 92°C, 2 min at 45°C and 2 min at 72°C. The Taq polymerase was obtained

from GIBCO-BRL and the concentrations of primers in the reaction mixture were 3  $\mu$ M for the first primer and 36  $\mu$ M for the second. After amplification of *S. citri* DNA, a single DNA fragment of the expected size (110bp) was obtained.

For homologous probes, i.e. of *S. citri* origin, hybridizations and washes were carried out under highly stringent conditions: hybridization at 45°C in a buffer containing 50% formamide and washes at 55°C with 0.1 $\times$ SSC (1 $\times$ SSC=150mM NaCl, 15mM Na<sub>3</sub>-citrate). For heterologous probes, i.e. from organisms different of *S. citri*, hybridizations and washes were carried out under less stringent conditions: hybridization at 37 to 42°C with 50% formamide, and washes at 37 to 40°C with 0.5 $\times$ SSC.

## RESULTS

### Separation of restriction fragments resulting from single digestions

As illustrated in Figure 1A and Table 2, digestion of the *S. citri* genome with restriction enzymes *Apa*I (target sequence: 5'-GGGCCC-3') or *Bss*HII (5'-GCGCGC-3') generated 9 fragments. They were labelled alphabetically from the largest to the smallest. Enzyme *Sal*I (5'-GTCGAC-3') generated 7 fragments. The presence of a small *Sal*I fragment (SIH) of 2.8 kbp (not visible upon PFGE) was demonstrated by running *Sal*I

**Table 1.** Gene markers used for the establishment of *S. citri* genomic map.

Probe number and locus name	Products	Source	References <sup>a</sup>
1: <i>rpsB-tsfx-spi-pfkA-pyk</i>	Ribosomal protein S2, EF-Ts, unknown protein, Spiralin, 6-Phosphofructokinase, Pyruvate kinase	<i>S. citri</i>	16
2: <i>rrn</i>	rRNA 16S	<i>S. citri</i>	17
3: <i>pyrG-purA-purB</i>	CTP synthetase, Adenylosuccinate synthetase and lyase	<i>S. citri</i>	
4: <i>fib</i>	Fibrillar protein	<i>S. citri</i>	18
5:SpV1-R8A2B DNA	RF of virus SpV1-R8A2B	<i>S. citri</i>	19
6:SpV1-S102 DNA	RF of virus SpV1-S102	<i>S. citri</i>	
7: <i>rpoD</i>	RNA polymerase $\sigma$ subunit	<i>S. citri</i>	
8: <i>trnB</i>	tRNA <sup>Val</sup> (UAC), tRNA <sup>Thr</sup> (UGU), tRNA <sup>Tyr</sup> (GUA), tRNA <sup>Gln</sup> (UUG), tRNA <sup>Lys</sup> (UUU), tRNA <sup>Leu</sup> (UAA)	<i>S. citri</i>	
9: <i>ileT-alaT</i>	tRNA <sup>Ile</sup> (GAU), tRNA <sup>Ala</sup> (UGC)	<i>S. citri</i>	
10: <i>trpT-trpU-serT</i> tRNA <sup>Ser</sup> (CGA)	tRNA <sup>Trp</sup> (UCA), tRNA <sup>Trp</sup> (CCA),	<i>S. citri</i>	
11: <i>trn</i>	Total tRNAs	<i>S. citri</i>	
12: <i>atp</i>	Membrane bound ATP synthase	<i>Mycoplasma</i> sp. PG50	20
13: <i>glyA</i>	$\alpha$ , $\beta$ , $\Gamma$ , $\delta$ subunits Serine hydroxymethyltransferase	<i>Mycoplasma</i> sp. PG50	20
14: <i>rpoB</i> , C	RNA polymerase $\beta$ , $\beta'$ subunits	<i>Mycoplasma</i> sp. PG50	20
15: <i>gyrA</i> , B	DNA gyrase A, B subunits	<i>M. pneumoniae</i>	21
16: <i>rpn</i>	Ribosomal proteins L2, S19, L22, S3, L16, L29, S17, L14, L24, L5, S14, S8, L6, L18, S5, L15, SecY, Adenylate kinase	<i>M. capricolum</i>	22
17: <i>kdp</i>	K <sup>+</sup> -stimulated ATPase	<i>E. coli</i>	23
18: <i>msr</i>	4.5S RNA	<i>M. mycoides</i>	24
19: <i>trnA</i>	tRNA <sup>Arg</sup> (ACG)-tRNA <sup>Pro</sup> (UGG)-tRNA <sup>Ala</sup> (UGC)-tRNA <sup>Met</sup> (CAU) tRNA <sup>Ile</sup> (CAU)-tRNA <sup>Ser</sup> (UGA) tRNA <sup>Met</sup> (CAU)-tRNA <sup>Asp</sup> (GUC) tRNA <sup>Phe</sup> (GAA).	<i>M. capricolum</i>	25
20: <i>argT</i>	tRNA <sup>Arg</sup> (UCU)	<i>M. capricolum</i>	25
21: <i>glyT</i>	tRNA <sup>Gly</sup> (UCC)	<i>M. capricolum</i>	25
22: <i>hisT</i>	tRNA <sup>His</sup> (GUG)	<i>M. capricolum</i>	25
23: <i>leuT</i>	tRNA <sup>Leu</sup> (CAA)	<i>M. capricolum</i>	25
24: <i>serT</i>	tRNA <sup>Ser</sup> (GCU)	<i>M. capricolum</i>	25

<sup>a</sup>Probes from references (20), (21), (22,25), and (24) have been kindly supplied by Christiansen, P.-C. Hu, A. Muto, and T. Samuelsson, respectively. Probes without a reference number are from our laboratory and are unpublished.

digested *S. citri* DNA onto a classical horizontal 0.7% agarose gel. As a result, 8 *SalI* sites were found in the *S. citri* genome.

Unlike *ApaI*, *BssHII* and *SalI*, the enzymes *SmaI* (5'-CCCGGG-3') and *SstII* (5'-GAGCTC-3') cleaved the *S. citri* genome more frequently, generating 17 and 14 fragments, respectively. In contrast, a single site for enzyme *NotI* (5'-GCGGCCGC-3') has been found in the *S. citri* genome, and is located within fragment ApE, BsB and SIa (Figure 1B).

The genome size of *S. citri* estimated from the digestion products is about 1780 kbp, a value similar to that obtained with the *NotI* linearised genome or the undigested linear molecule in TAFE gels.

### Ordering of the restriction fragments and physical mapping of the *S. citri* genome

For ordering of the restriction fragments, reciprocal double digestions of *S. citri* DNA by pairs of enzymes were carried out. Most of the different restriction fragments were successfully connected to each other as indicated in Figure 2. For some sites such as ApF, E, H and I which are all within the large fragments SIa or BsB, the second digestion with *ApaI* was not informative. Thus, they were mapped as follows. The *S. citri* DNA was firstly cleaved to completion with *SalI* or *BssHII*, then partially digested with *ApaI* (10 min at 37°C). After gel electrophoresis and transfer, the membrane was hybridized with probe 4 (containing the fibrillar protein gene). Since this gene is located at one end of fragments SIa or BsB (see below), all the fragments resulting from partial digestion with *ApaI* and hybridizing with this probe will give the order of internal *ApaI* sites (figure 2). The order of these fragments was shown to be: ApG-H-E-F-I-A.

In addition, some small fragments such as BsI, SIG and H, SmP and Q as well as SsL, M and N were localized on the map by hybridization. These fragments were excised from TAFE gels, purified and used as probes for hybridizations with membranes containing products of single restriction digests. Fragment SIG is between SID and A, SmP is located between SmA and H, SmQ is between SmB and I, SsL is between SsF and C, SsM is between

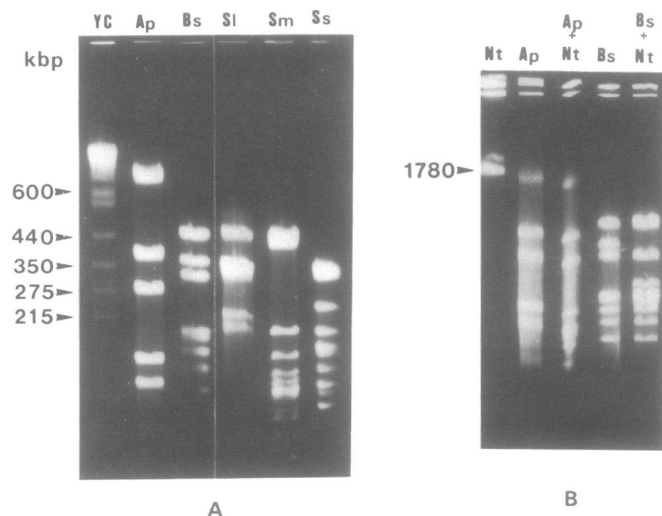
SsB and G, SsN is between SsA and D, and BsI is located upstream of BsA.

Finally the smallest *SalI* fragment, SIH, gave the same hybridization pattern than that obtained with fragment SIG,

**Table 2.** Data on restriction fragments from the DNA of *S. citri* R8A2HP

Enzyme	Fragment	Size (kbp)	Hybridization with probes <sup>a</sup>
<i>ApaI</i>	ApA	676	1,4,5,6
	ApB	369	5,9,10,11,15,19,23
	ApC	270	3,5,12,13,14
	ApD	121	5,6,9,11,19
	ApE	121	7,11,20
	ApF	88	2,8,11,17,18,21,22,24
	ApG	86	5,11,16,19
	ApH	28	
	ApI	27	
	TOTAL	1786	
<i>BssHII</i>	BsA	449	5,10,11,12,13,14,15
	BsB	357	2,4,7,8,11,16,17,18,20,21,22,24
	BsC	310	5
	BsD	177	5,6,9,11,19
	BsE	167	5
	BsF	136	3,5,11,19
	BsG	106	1,5,6
	BsH	73	5
	BsI	8	11,23
	TOTAL	1783	
<i>SalI</i>	SIa	437	2,4,5,7,8,11,16,17,18,19,20,21,22,24
	SIB	329	5,9,10,11,19,23
	SIC	320	5,6
	SID	304	3,5,12,13,14,15
	SIE	201	1,5
	SIF	176	5,6,9,11,19
	SIG	10	
	SIH	3	
	TOTAL	1781	
<i>SmaI</i>	SmA	432	5,10,11,12,13,14,15
	SmB	420	1,5,6
	SmC	170	5,6,9,11,19
	SmD	125	4,17
	SmE	98	2,7,8,11,18,20,21,22,24
	SmF	82	5,11,16,19
	SmG	75	5
	SmH	72	3,5
	SmI	70	5
	SmJ	60	5,11,23
	SmK	45	5,11,19
	SmL	38	
	SmM	38	
	SmN	25	
	SmO	25	
	SmP	5	
	SmQ	3	
	TOTAL	1783	
<i>SstII</i>	SsA	340	5,6,9,11,19
	SsB	283	5,12,13,14,15
	SsC	230	1,5,6
	SsD	135	5,10,11,23
	SsE	135	5,16
	SsF	130	5
	SsG	120	3,5,11,19
	SsH	100	2,8,11,17,18,21,22,24
	SsI	80	4
	SsJ	68	5,10,11
	SsK	45	7,11,20
	SsL	42	5
	SsM	12	
	SsN	12	
	TOTAL	1782	

a: probe number is as in table 1.



**Fig. 1.** Pulsed field gel electrophoresis of restricted genomic DNA from *S. citri* strain R8A2HP. A: single digestions; B: single and double digestions; Yc: yeast chromosomes; Ap: *ApaI*; Bs: *BssHII*; SI: *SalI*; Sm: *SmaI*; Ss: *SstII*; Nt: *NotI*.

suggesting linkage of these two fragments on the map. This has been confirmed by screening a lambda phage library of *S. citri* DNA and finding a recombinant phage which bears the majority of fragment SIG, the entire SIH as well as a small part of SIA (data not shown).

Compilation of the above results led to the construction of the physical map of the *S. citri* genome presented in figure 2. One end of the largest restriction fragment, ApA, is taken arbitrarily as the starting point (0 on the map).

### Localization of some genetic markers on the *S. citri* genome

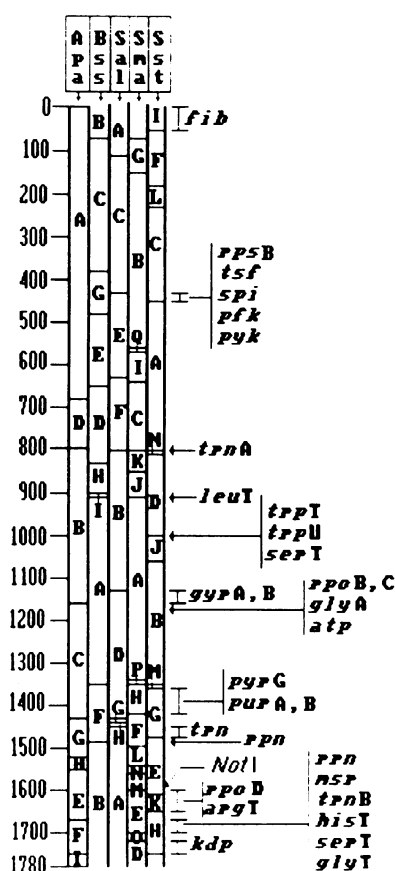
**Protein-encoding genes.** Three sets of *S. citri* protein-encoding genes have been cloned and sequenced so far: the fibrillar protein gene (*fib*), the spiralin (*spi*) and adjacent genes (*rpsB*, *tsf*, *pfk* and *pyk*) and a cluster of three genes involved in nucleotide metabolism: *pyrG-purA-purB*. These were located onto the map by hybridization experiments as exemplified in figure 3. In each case, a single fragment for each restriction enzyme hybridized with the probe, suggesting that there is only one copy of these genes in the genome. However the orientation of the transcription remains to be determined.

The potential location of *rpoD*, encoding the general Sigma factor, was mapped using an amplified DNA fragment as a probe (see materials and methods). We are currently verifying that it

is indeed *rpoD* by sequencing a recombinant plasmid screened with the same probe.

All the other protein-encoding genes that were mapped on the *S. citri* genome were from heterologous origins but they were supposed to be conserved because of their functions (transcription, replication, translation, metabolism) and most of them came from *Mycoplasma* spp. phylogenetically close to *S. citri*. Used as probes, they gave weak but clear cut hybridization signals, suggesting that the putative *loci* for these genes existed in the *S. citri* genome. The ribosomal protein gene cluster from *M. capricolum* (*rpn*) was found to be localized in a single locus on the *S. citri* genome. The fact that the overlap between fragments SsE and SIA is rather small (25 kbp) while the probe is rather large (8 kbp) may suggest that the content of this cluster is common between the two organisms. Similarly, the genes, from *Mycoplasma* sp. PG50, encoding *rpoB*, *rpoC*, *glyA* and *atpH*, A, G, D (all located within the same 23 kbp DNA fragment in *Mycoplasma* sp. PG50)(20) were found to hybridize to the same fragments of *S. citri* DNA. In screening a lambda phage genomic library made with *S. citri* DNA partially cleaved by *MboI*, we found two clones hybridizing with both *rpoB,C* and *glyA* (data not shown). These clones possess an *Apal* site, indicating that *rpoB,C* and *glyA* are located at the beginning of fragment ApC. If the organisation of these genes is conserved between *Mycoplasma* sp. PG50 and *S. citri*, as it is with *M. capricolum* and *M. mycoides*, *atp* genes should be located nearby. *GyrA* and *gyrB* are found in the vicinity of these genes, since they have to be located in the small overlapping region of fragments ApB and SID.

In *E. coli* and other bacteria, two distinct sets of ATPases exist: membrane bound ATP synthase encoded by several *atp* genes and K<sup>+</sup>-stimulated ATPase encoded by *kdp*. Because it has been previously suggested that *S. citri* also has two sets of ATPases (30), we used the appropriate probes to map the potential genes. As mentioned above four of the *atp* genes were identified in the vicinity of *rpoB* and C. Using the *kdp* gene from *E. coli*, we obtained an hybridization in a region located far away from the *atp* operon. This result is in agreement with the proposal by Simoneau and Labarère (30), that *S. citri* possesses two distinct ATPases but has to be confirmed by sequence data.



**rRNA and tRNA genes.** It has been previously shown that *S. citri* possesses only one ribosomal RNA operon (*rrn*) in which the genes are arranged in the classical eubacterial order: 5'-16S-23S-5S-3' (31). In contrast with several bacterial and mollicute species, no tRNA genes were found in the intergenic regions and/or in the promoter upstream sequences (32). Use of the 16S rDNA as a probe resulted in a single hybridizing fragment for each restriction enzyme, and made it possible to locate *rrn* within the overlap of ApF and SmE.

In *M. capricolum*, a mycoplasma phylogenetically related to *S. citri*, there are 30 tRNA genes encoding 29 tRNA species (25). Recent studies from our laboratory suggest a similar situation for *S. citri* (C. Citti, unpublished). Consequently, we used *M. capricolum* cloned tRNA genes, *S. citri* total tRNAs and several *S. citri* cloned tRNA genes, as probes, to position tRNA loci onto the *S. citri* map. The results are summarized in table 3 and illustrated in figure 4. Hybridizations between membranes from TAFE gels and total tRNAs from *S. citri* indicated that tRNA genes are distributed in 6 areas. Three *S. citri* tRNA gene clusters have been identified, cloned and partially sequenced (C. Citti, unpublished). A cluster of 6 tRNA genes (*trnB*), encoding tRNA<sup>Val</sup>(UAC)-tRNA<sup>Thr</sup>(UGU)-tRNA<sup>Tyr</sup>(GUA)-tRNA<sup>Gln</sup>(UUG)-tRNA<sup>Lys</sup>(UUU)-tRNA<sup>Leu</sup>(UAA), is closed to the *rrn* operon. A DNA fragment, containing a three tRNA gene cluster, encoding tRNA<sup>Trp</sup>(UCA)-tRNA<sup>Trp</sup>(CCA)-tRNA<sup>Ser</sup>(CGA), is located

within the overlap between BsA and SmA. Because it hybridizes with both fragments SsD and SsJ, it contains a *SsrII* site and can be precisely placed on the junction of these two fragments. A third DNA fragment containing a two genes cluster, encoding tRNA<sup>Ile</sup>(GAU)-tRNA<sup>Ala</sup>(UGC), hybridized with fragments ApD and B, BsD, SIF and B, and SmC. Because the sequence of tRNA<sup>Ile</sup>(GAU) contains sites for *Apal* and *SalI*, this cluster is located precisely at the junction of ApD and B.

Thus, genes for 10 tRNA species of *S. citri* were identified and placed on the map. The remaining tRNA genes were tentatively localised using *M. capricolum* cloned tRNA genes as probes (25). A 9 genes cluster (*trnA*) composed of tRNA<sup>Arg</sup>(A-CG)-tRNA<sup>Pro</sup>(UGG)-tRNA<sup>Ala</sup>(UGC)-tRNA<sup>Met</sup>(CAU)-tRNA<sup>Ile</sup>(CAU)-tRNA<sup>Ser</sup>(UGA)-tRNA<sup>fMet</sup>(CAU)-tRNA<sup>Asp</sup>(GUC)-tRNA<sup>Phe</sup>(GAA), was found to hybridize strongly with the same fragments than the *S. citri* tRNA<sup>Ile</sup>(GAU)-tRNA<sup>Ala</sup>(UGC) cluster, but also with fragment SmK. The fact that *trnA* contains tRNA<sup>Ala</sup>(UGC) explains this result but reaction with SmK fragment indicated that other tRNA genes, homologous to those in *trnA*, are located in the vicinity of this two genes cluster. Because a similar 10 genes cluster, containing the 9 tRNA genes of *trnA*, was found in *S. melliferum* (33) it is expected that such a cluster is located in this region of the *S. citri* chromosome. In addition *trnA* hybridizes weakly with SmF, showing that this cluster may be split between two different areas in the *S. citri* genome. *LeuT*, encoding tRNA<sup>Leu</sup>(CAA), is located within the 8 kbp *BsI* fragment. *ArgT*, tRNA<sup>Arg</sup>(UCU), *glyT*, tRNA<sup>Gly</sup>(UCC), and *serT*, tRNA<sup>Ser</sup>(GCU), are located in the vicinity of *rrn* and of *trnB*. All the remaining tRNA genes have not been positioned because no clear hybridization was obtained with probes of *M. capricolum* origin. Nevertheless, these genes have to belong to the regions identified with total tRNAs as a probe.

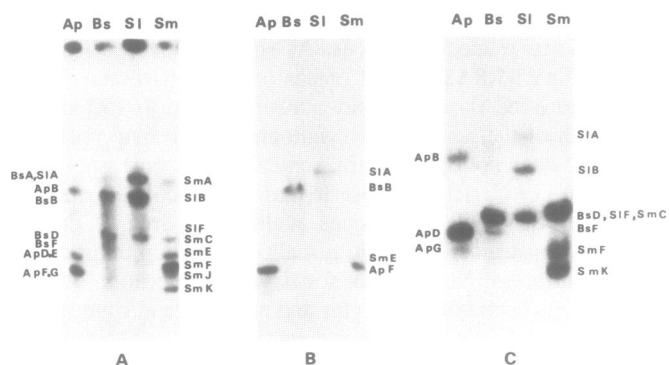
Recently, a minor small RNA, *msr*, showing some similarity with the 4.5S RNA of *E. coli*, has been identified in both *M. mycoides* and *M. capricolum* (24,25). The cloned *M. mycoides* *msr* gene was used as probe, and found to hybridize to fragments ApF, BsB, SIA and SmE, and, consequently, to be located in the vicinity of *rrn*.

**Table 3.** Location of tRNA genes on *S. citri* physical map

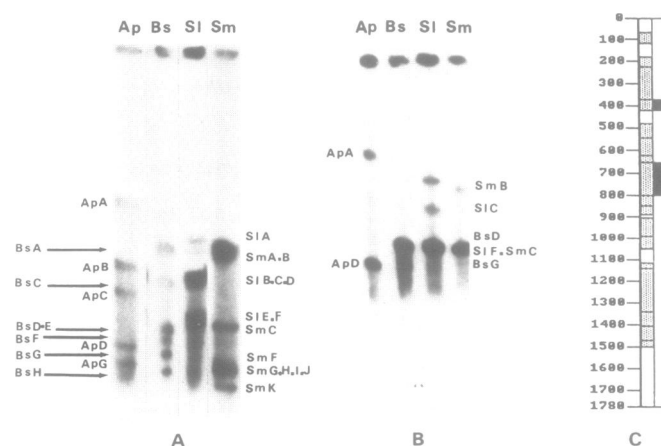
Probe <sup>a</sup>	Restriction fragments hybridizing with the probe				
	<i>Apal</i>	<i>BssHII</i>	<i>SalI</i>	<i>SmaI</i>	<i>SsrII</i>
<i>trn</i> (11)	B,D,E, F,G	A,B,D,F I <sup>b</sup>	A,B,F	A,C,E, F,J,K	A,D,G, H,J,K
<i>trnB</i> (8)	F	B	A	E	H
<i>ileT-alaT</i> (9)	B,D	D	B,F	C	A
<i>trpT,U</i> (10)	B	A	B	A	D,J
<i>trnA</i> (19)	B,D,G	D,F	A,B,F	C,F,K	A,G
<i>argT</i> (20)	E	B	A	E	K
<i>glyT</i> (21)	F	B	A	E	H
<i>hisT</i> (22)	F	B	A	E	H
<i>leuT</i> (23)	B	I	B	J	D
<i>serT</i> (24)	F	B	A	E	H

a: the number in parenthesis refers to the probe number in table 1.

b: this fragment, not seen in figure 4A, is revealed upon longer exposure.



**Fig. 4.** Hybridizations with probes representative of tRNA genes. A: total tRNAs purified from *S. citri*; B: *trnB*; C: *trnA*; enzyme abbreviations are as in figure 1.



**Fig. 5.** Hybridizations of *S. citri* genomic fragment with SpV1-DNA used as probe. A: SpV1-R8A2B; B: SpV1-S102; enzyme abbreviations are as in figure 1.; C: Distribution of SpV1-related sequences on the *S. citri* genome. Each delimited area contains at least one copy of viral-related sequences. Shaded areas: SpV1-R8A2B, black areas: SpV1-S102.

**SpV1-related sequences.** Most of the *S. citri* strains, isolated from either plants or insects, are infected by a filamentous phage, SpV1, with a single stranded DNA genome of 8 kb (19). The first such virus was isolated from strain R8A2B (SpV1-R8A2B) and it was demonstrated that all the *S. citri* strains contain, in their genomes, several sequences that hybridize with the viral DNA, indicating that SpV1-related sequences may be present in different regions of the genome (34). We used the replicative form (RF) of SpV1-R8A2B, as a probe, to evaluate the number and the distribution of such sequences on the *S. citri* map. As shown in figure 5A, several fragments for each restriction enzyme were revealed and one could deduce that there were at least 7 insertion sites with viral-like sequences within the genome. Further experiments, involving double digestions, indicated that, indeed, the minimal number of insertion sites was 17. In order to test the possibility that genomic sequences hybridizing with the RF represent only portions of the viral genome we used two different fragments of the RF as probes. They gave the same hybridization pattern as the entire RF did, suggesting that viral-like sequences are made of full-length viral genomes (data not shown). Recently, a second viral isolate has been obtained from *S. citri* strain S102 (SpV1-S102). While resembling the SpV1-R8A2B virus in morphology and genome characteristics, no sequence homology was found between the two SpV1-type viruses (J. Renaudin, unpublished). After hybridization of SpV1-S102 DNA with restricted *S. citri* genomic DNA, it was found that two regions of the genome gave hybridization signals; in one region the signal was strong, but weak in the other (figure 5B.). Thus, the genome of *S. citri*, strain R8A2HP, contains two sets of SpV1-S102 related sequences, one being longer or more homologous to the probe than the other. In consequence, this brings to 19 the minimal number of viral-related sequences in the genome which could account for up to 150 kbp of DNA sequences, in other words, one 12th of the whole genome. Interestingly enough, a region of 350 kbp, from BsB to SmG,

does not contain any SpV1-related sequences (figure 5C.). It represents one of the two G+C rich region and also contains *rm*, *rpoD*, several tRNA genes and *fib*.

## DISCUSSION

### Restriction site frequency

Restriction sites containing only G and C nucleotides (Ap, Bs, Sm and St) are not uniformly distributed on the genome. Indeed they cluster in certain areas. For example one half of these sites are localized in less than one third of the genome (map positions 1300 to 1780). Furthermore a small region (from 780 to 900) is extremely rich in such sites (10 sites within 120 kbp). These results indicate that the G+C content is variable along the genome. This non-uniform distribution of G+C bases has been previously observed for *M. mycoides* and *M. capricolum* genomes (7,12). Because the *S. citri* genome contains 26% G+C, the predicted frequency of a site such *ApaI* (5'-GGGCCC-3') is  $[0.13]^6 = 1/200,000$  bp, when calculated from mononucleotide frequencies. Thus 9 sites for *ApaI*, *BssHII*, *SmaI* or *SstII* should be present in the genome. This is the case for *ApaI* and *BssHII* but the real number of *SmaI* and *SstII* sites is larger than the predicted one. The G+C rich regions contain several essential, conserved, genes such as those encoding rRNAs, tRNAs and ribosomal proteins.

The *S. citri* genome contains 8 *SalI* sites. This low number is somewhat surprising since the predicted frequency of this site (5'-GTTCGAC-3') is  $[0.13]^4 \cdot [0.37]^2 = 1/25,000$  bp i.e. 10 fold higher than the observed one. We looked for the possibility that one or several di- or trinucleotides which are part of this site could be underrepresented within the genome. This was done by di- and trinucleotide frequency analysis in the 23,844 nucleotides sequenced so far. No significant difference between observed and predicted frequencies was observed and this could not account for the low number of *SalI* sites. Thus, another explanation is to suppose that the majority of *SalI* sites are part of a sequence which is methylated within the *S. citri* genome. Indeed the 16S rRNA gene sequence has been established (32; accession number:X63781) and it contains one *SalI* site which is not revealed upon Southern blot analysis. Interestingly enough this low frequency of *SalI* sites has also been observed for *M. mycoides*, *M. capricolum* and relatives (35), indicating that a similar methylation system may exist in these organisms.

### SpV1-related sequences

A striking feature of the *S. citri* genome is the ubiquitous presence of SpV1 virus-related sequences. As shown above, at least 17 copies of SpV1-R8A2B and 2 copies of SpV1-S102 have been found (figure 5C.) and this may account for up to 150 kbp or one twelfth of the genome. A single area, from map positions 1500 to 70 does not contain viral sequences and corresponds to one of the G+C rich regions. It may be speculated that the absence of SpV1-like sequences within this area is due to functional constraints which do not allow any disruption of the gene organization. No evidence for *de novo* integration of viral sequences has been obtained so far and no biological significance for these sequences has been suggested. Nevertheless, the presence of such a large number of viral-like sequences speaks in favour of a biological role. These viral-like sequences may be considered as repeated elements and, as such, may play a role in genomic rearrangements.

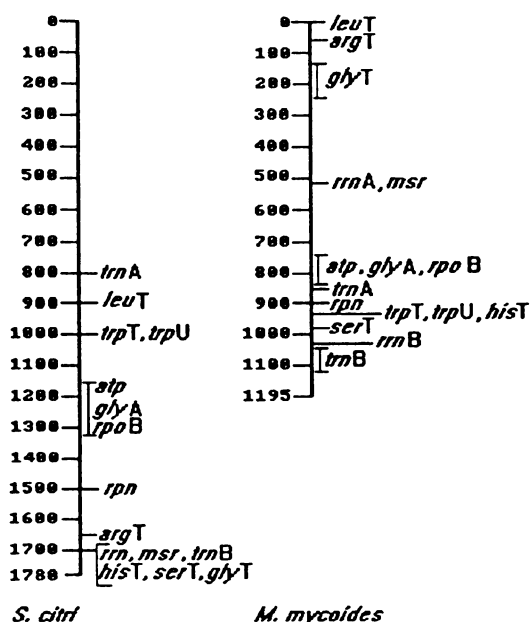


Fig. 6. Comparison of the genomic maps of *S. citri* and *M. mycoides* subsp. *mycoides*, strain Y goat.

### Comparison with *M. mycoides* genome

One goal of the present work was to compare the genomic organization of *S. citri* with that of other Mollicutes and, especially *M. mycoides* which belongs to the same phylogenetic group than *S. citri*, as determined by 16S rRNA sequence comparisons (7). At first sight, the two genomes show striking differences which reside not only in the genome size (1780 kbp versus 1195 kbp) but also in restriction site distribution and genetic organization, especially for tRNA genes (figure 6.). Nevertheless comparisons between *S. citri* and *M. mycoides* subsp. *mycoides*, strain Y-goat, reveal several common features (figure 6.). For example, in both genomes, *rpoB* & *C*, *glyA* and *atp* are located together. Because it was shown that the origin of replication of *M. mycoides* was in the vicinity of these genes (8) it may be postulated that this is also the case for *S. citri*. The presence of *gyrA* & *B* in the same area supports this hypothesis since these genes, as well as *rpoB* & *C*, are close to the origin in *Bacillus subtilis* (36). In *E. coli*, *gyrB* and *atp* are near the origin while *rpoB* & *C* are located 6 minutes away (37).

In *M. mycoides* there are two *rrn* loci located far away from each other; *rrnA* is close to *msr* and has two tRNA genes in its 5' upstream sequences, while *rrnB* is close to *trnB*. In the *S. citri* genome, the single *rrn* locus has no tRNA genes in its immediate flanking sequences and is located in the vicinity of *msr*, and *trnB*. This may suggest that one of the event leading from the common ancestor of these two organisms to *S. citri* could be a recombination between the two *rrn* loci with the subsequent loss of one *rrn* operon and grouping of adjacent genes.

It has been suggested that Mollicutes are fast evolving organisms with a high mutation rate (38). In addition, from recent work on genomic maps within the *M. mycoides* cluster, it seems that these organisms undergo significant genomic rearrangements from one strain to the other (35). We are currently addressing this question in establishing the genomic maps for different strains of *S. citri* and for other *Spiroplasma* species belonging to the same serogroup. In addition, this may bring information regarding the distribution of viral-related sequences and their potential role in rearrangements. Finally, from comparison between *M. mycoides* and *S. citri* it seems that G+C poor regions may be species-specific, since few conserved genes are found within these areas. Studies of additional species within the same phylogenetic group will help in assessing this hypothesis.

### REFERENCES

- Pyle, L.E., Corcoran, L.N., Cocks, B.J., Bergemann, A.D., Whitley, J.C. and Finch, L.R. (1988). Nucleic Acids Res., **16**, 6015–6025.
- Neimark, H.C. and Lange, C.S. (1990). Nucleic Acids Res., **18**, 5443–5448.
- Robertson, J.A., Pyle, L.E., Stemke, G.W. and Finch, L.R. (1990). Nucleic Acids Res., **18**, 1451–1455.
- Grau, O., Laigret, F., Carle, P., Tully, J.G., Rose, D.L. and Bové, J.M. (1991). Int. J. Syst. Bacteriol., **41**, 473–478.
- Su, C.J. and Baseman, J.B. (1990) J. Bacteriol., **172**, 4705–4707.
- Colman, S.D., Hu, P.-C., Litaker, W., and Bott, K.F. (1990) Mol. Microbiol., **4**, 683–687.
- Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T. G., Van Etten, J., Maniloff, J. and Woese, C.R. (1989). J. Bacteriol., **171**, 6455–6467.
- Pyle, L.E. and Finch, L.R. (1988). Nucleic Acids Res., **16**, 6027–6040.
- Wenzel, R. and Herrmann, R. (1988). Nucleic Acids Res., **16**, 8323–8336.
- Bautsch, W. (1988). Nucleic Acids Res., **16**, 11461–11468.
- Cocks, B.G., Pyle, L.E. and Finch, L.R. (1989). Nucleic Acids Res., **17**, 6713–6719.
- Krause, D.C. and Mawn, C.B. (1990). J. Bacteriol., **172**, 4790–4797.
- Pyle, L.E., Taylor, T. and Finch, L.R. (1990) J. Bacteriol., **172**, 7265–7268.
- Whitcomb, R.F. (1983). In Razin, S. and Tully, J.G. (eds.), Methods in mycoplasmaology. Academic Press, New York, pp. 147–159.
- Pyle, L.E. and Finch, L.R. (1988). Nucleic Acids Res., **16**, 226–2268.
- Chevalier, C., Saillard, C. and Bové J.M. (1990). J. Bacteriol., **172**, 2693–2703.
- Laigret, F., Grau, O. and Bové J.M. (1990). In Stanek, G., Cassell, G.H., Tully, J.G. and Whitcomb, R.F. (eds.), Recent advances in Mycoplasmaology. Zbl. Bakteriologie. Gustav Fischer Verlag, Stuttgart, New York, pp. 435–440.
- Williamson, D.L., Renaudin, J. and Bové, J.M. (1991). J. Bacteriol., **173**, 4353–4362.
- Renaudin, J., Aullo, P., Vignault, J.C. and Bové, J.M. (1990). Nucleic Acids Res., **18**, 1293.
- Rasmussen, O.F. and Christiansen, C. (1990). In Stanek, G., Cassell, G.H., Tully, J.G. and Whitcomb, R.F. (eds.), Recent advances in Mycoplasmaology. Zbl. Bakteriologie. Gustav Fischer Verlag, Stuttgart, New York, pp. 315–323.
- Colman, S.D., Hu, P.-C. and Bott, K.F. (1990). Mol. Microbiol., **4**, 1129–1134.
- Ohkubo, S., Muto, A., Kawachi, Y., Yamao, F. and Osawa, S. (1987). Mol. Gen. Genet., **210**, 314–322.
- Hesse, J.E., Wiczorek, L., Altendorf, K., Reicin, A., Dorus, E. and Epstein, W. (1984). Proc. Natl. Acad. Sci. USA., **81**, 4746–4750.
- Samuelsson, T. and Guindy, Y. (1990). Nucleic Acids Res., **18**, 4938.
- Muto, A., Andachi, Y., Yuzawa, H., Yamao, F. and Osawa, S. (1990). Nucleic Acids Res., **18**, 5037–5043.
- Southern, E.M. (1975). J. Mol. Biol., **98**, 503–517.
- Reed, K.C. and Mann, D.A. (1985). Nucleic Acids Res., **13**, 7207–7221.
- Zubay, G. (1962). J. Mol. Biol., **4**, 347–356.
- Engel, J.N. and Ganem, D. (1990). J. Bacteriol., **172**, 2447–2455.
- Simoneau, P. and Labarère, J. (1991). J. Gen. Microbiol., **137**, 179–185.
- Grau, O., Laigret, F. and Bové, J.M. (1990). In Stanek, G., Cassell, G.H., Tully, J.G. and Whitcomb, R.F. (eds.), Recent Advances in Mycoplasmaology. Zbl. Bakteriologie. Gustav Fischer Verlag, Stuttgart, New York, pp. 895–897.
- Grau, O. (1991). PhD thesis, Université de Bordeaux II, France.
- Rogers, M.J., Steinmetz, A.A. and Walker, R.T. (1986). Nucleic Acids Res., **7**, 3145.
- Renaudin, J., Bodin-Ramiro, C., Vignault, J.C. and Bové, J.M. (1990). In Stanek, G., Cassell, G.H., Tully, J.G. and Whitcomb, R.F. (eds.), Recent advances in Mycoplasmaology. Zbl. Bakteriologie. Gustav Fischer Verlag, Stuttgart, New York, pp. 125–130.
- Whitley, J.C. (1991). PhD thesis, University of Melbourne, Australia.
- Piggot, P.J., Amjad, M., Wu, J.J., Sandoval, H. and Castro, J. (1990). In Harwood, C.R. and Cutting, S.M. (eds.), Molecular biological methods for Bacillus. John Wiley and Sons, New York, pp. 439–450.
- Bachmann, B.J. (1990). Microbiol. Rev., **54**, 130–195.
- Woese, C.R., Stackebrandt, E. and Ludwig, W. (1985). J. Mol. Evol., **21**, 305–316.